PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : H04N 17/00

A1

(11) International Publication Number: WO 00/01165

(43) International Publication Date: 6 January 2000 (06.01.00)

(21) International Application Number: PCT/GB99/01894

(22) International Filing Date: 15 June 1999 (15.06.99)

(30) Priority Data:
98305107.9 29 June 1998 (29.06.98) EP

(71) Applicant *(for all designated States except US)*: BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

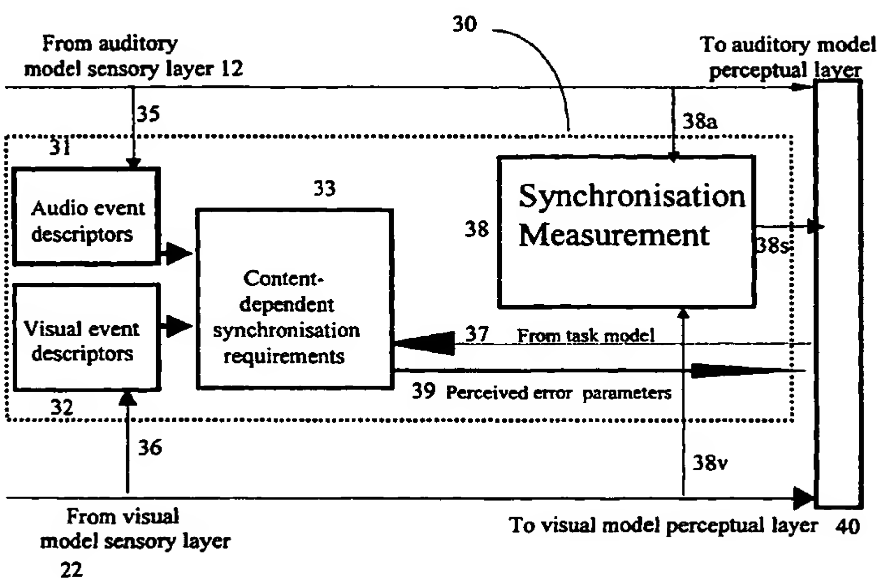
(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: HOLLIER, Michael, Peter [GB/GB]; 15 Herbert Road, Grange Farm, Kesgrave, Ipswich, Suffolk IP5 2XX (GB).

(74) Agent: LIDBETTER, Timothy, Guy, Edwin; BT Group Legal Services, Intellectual Property Dept., Holborn Centre, 8th floor, 120 Holborn, London EC1N 2TE (GB).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: MEASUREMENT OF PERFORMANCE OF COMMUNICATIONS SYSTEMS



(57) Abstract

The subjective quality of an audio–visual stimulus is measured by measuring the actual synchronisation errors 38s between the audio and visual elements (12, 22) of the stimulus, identifying characteristics of audio and visual cues in the stimulus (31), and generating a measure of subjective quality from said errors and characteristics. The nature of the cue has an effect on the perceptual significance of a given value of synchronisation error, and this can be used to relax tolerances to such errors when appropriate, or to provide an accurate measure of the quality of the signal as it would be perceived by a human subject.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

MEASUREMENT OF PERFORMANCE OF COMMUNICATIONS SYSTEMS

This invention relates to signal processing. It is of application to the testing of communications systems and installations, and to other uses as will be described. The term "communications system" covers telephone or television networks and equipment, public address systems, computer interfaces, and the like.

It is desirable to use objective, repeatable, performance metrics to assess the acceptability of performance at the design, commissioning, and monitoring stages of communications services provision. However, a key aspect of system performance is the measurement of subjective quality, which is central in determining customer satisfaction with products and services. The complexity of modern communications and broadcast systems, and in particular the use of data reduction techniques, renders conventional engineering metrics inadequate for the reliable prediction of perceived performance. Subjective testing using human observers is expensive, time consuming and often impractical, particularly for field use. Objective assessment of the perceived (subjective) performance of complex systems has been enabled by the development of a new generation of measurement techniques, which emulate the properties of the human senses. For example, a poor value of an objective measure such as signal-to-noise performance may result from an inaudible distortion. A model of the masking that occurs in hearing is capable of distinguishing between audible and inaudible distortions.

The use of models of the human senses to provide improved understanding of subjective performance is known as *perceptual modelling.* The present applicants have a series of previous patent applications referring to perceptual models, and test signals suitable for non-linear speech systems, including WO 94/00922, WO 95/01011 and WO 95/15035.

To determine the subjective relevance of errors in audio systems, and particularly speech systems, assessment algorithms have been developed based on models of human hearing. The prediction of audible differences between a degraded signal and a reference signal can be thought of as the *sensory layer* of a perceptual analysis, while the subsequent categorisation of audible errors according to their subjective effect on overall signal quality can be thought of as the *perceptual layer*.

An approach similar to this auditory perceptual model has also been adopted for a visual perceptual model. In this case the *sensory layer* reproduces the gross psychophysics of the sensory mechanisms, in particular spatio-temporal sensitivity, (known as the human visual filter), and masking due to spatial frequency, orientation and temporal frequency.

A number of visual perceptual models are under development and several have been proposed in the literature.

The subjective performance of multi-modal systems depends not only on the quality of the individual audio and video components, but also on interactions between them. Such effects include "quality mis-match", in which the quality presented in one modality influences perception in another modality. This effect increases with the quality mis-match.

The information content of the signal is also important. This is related to the task undertaken but can vary during the task. For present purposes, "content" refers to the nature of the audio/visual material during any given part of the task.

The type of task or activity undertaken also has a substantial effect on perceived performance. As a simple example, if the video component dominates for a given task then errors in the video part will be of greatest significance. At the same time audio errors which have high attentional salience (are "attention grabbing") will also become important. The nature of the task undertaken influences the split of attention between the modalities, although this may also vary more randomly if the task is undemanding.

However, important though these factors are, they are in general difficult to define, and to use for making objective measurements. Nevertheless, the inventor has identified some cross-modal effects which can be derived from objective measurements.

According to the invention there is provided a method of determining the subjective quality of an audio-visual stimulus, comprising the steps of:

measuring the actual synchronisation errors between the audio and visual elements of the stimulus,

identifying characteristics of audio and visual cues in the stimulus,

and generating a measure of subjective quality from said errors and characteristics.

According to another aspect there is provided apparatus for determining the subjective quality of an audio-visual stimulus, comprising means for measuring the actual synchronisation errors between the audio and visual elements of the stimulus, means for the identification of characteristics of audio and visual cues in the stimulus, and means for generating a measure of subjective quality from said synchronisation errors and characteristics.

It has been observed experimentally that human subjects have different sensitivities to a given synchronisation error, depending on the type of cue with which it is associated. Thus, poorly-synchronised stimuli containing certain cue types will be perceived as of lower quality than equally poorly-synchronised stimuli containing other cue types. Synchronisation tolerances have been an essential consideration in television broadcasting for many years. However, for emerging telepresence technologies, synchronisation must be dynamically controlled. Audio/video synchronisation error detection is dependent on the task undertaken, the nature of the stimulus (content) and whether the error results in the audio leading or lagging the video *[ITU-T Recommendation J.100, "Tolerances for transmission time differences between vision and sound components of a television signal", 1990*].

Results to be presented later in this specification illustrate that the synchronisation tolerances can be relaxed for certain types of content and that the subjectivity of synchronisation error remains relatively low over a much greater range of values for these types.

Although, in general, information content is not measurable by an objective test, certain cue types have been identified on which human sensitivity to synchronisation error depends, and which are distinguishable by such tests.

Synchronisation errors are also relatively easy to measure, so the invention allows a network operator to determine by objective measures whether the current error is perceptually significant, having regard to the nature of the cue.

The characteristics of the audio and visual cues are preferably used to generate one or more synchronisation error tolerance values, which may correspond to different degrees of perceptual error, for example as measured by human subjects. The audio-visual stimulus can be monitored for occurrences of synchronisation errors exceeding such tolerance values to provide a quantitative output. The means generating the stimulus may be controlled dynamically to

maintain the synchronisation in a predetermined relationship with the said tolerance values, for example by buffering the earlier-arriving stimulus or by omitting elements from the later-arriving one to bring them into synchronism. Maintenance of synchronisation can make considerable demands on an audio-visual system. Buffering requires a memory capacity. Alternatively, if channels are congested, data packets of one or other channel (sound or vision) may have to be sacrificed to maintain synchronisation to a given level, reducing the signal quality of that channel. Therefore, if a more relaxed tolerance level can be applied at certain times, greater synchronisation errors can be allowed, thereby reducing the required channel capacity and/or the amount of lost data.

Where there are several channels in use, each carrying different stimulus types, they may be controlled such that they all have the same perceptual quality value, although the synchronisation errors themselves may be different.

A further application of the invention is in the real-time generation of audio-visual events in virtual environments, in particular real-time optimisation of synthetic people such as animated talking faces. The process may be used in particular for the matching of synthetic head visemes (mouth shape) transitions with the acoustic waveform data generating the speech to be represented, thereby generating ore realistic avatars.

An embodiment of the invention will now be described, by way of example only, with reference to the Figures.

Figure 1 shows in schematic form the principal components of a multi-sensory perceptual measurement system.

Figure 2 shows a synchronisation perception measurement component of the system of Figure 1.

Figures 3 and 4 illustrate experimental data indicative of the behaviour modelled by the synchronisation measurement component.

Figure 5 illustrates schematically the use of the system of Figure 2 in generating visemes for an avatar.

A suitable architecture for a multi-sensory model is shown in Figure 1. The main components are:

- auditory and visual sensory models 10, 20;
- a cross-modal model 30, The cross-modal model 30 includes a synchronisation perceptual model shown in detail in Figure 2

∘ a scenario-specific perceptual layer 40.

An auditory sensory layer model component 10 comprises an input 11 for the audio stimulus, which is provided to an auditory sensory layer model 12. The auditory model 12 measures the perceptual importance of the various auditory bands and time elements of the stimulus, and generates an output 16 representative of the audible error as a function of auditory band (pitch) and time. This audible error may be derived by comparison of the perceptually modified audio stimulus 13 and a reference signal 14, the difference being determined by a subtraction unit 15 to provide an output 16 in the form of a matrix of subjective error as a function of auditory band and time, defined by a series of coefficients E_{da1}, E_{da2}, ..., E_{dan}. Alternatively the model may produce the output 16 without the use of a reference signal, for example according to the method described in international patent specification number WO96/06496.

A similar process takes place with respect to the visual sensory layer model 20. However, in this context a further step is required. The image generated by the visual sensory layer model 22 is analysed in an image decomposition unit 27 to identify elements in which errors are particularly significant, and weighted accordingly, as described in international patent specification number WO97/32428. This provides a weighting function for those elements of the image which are perceptually the most important. In particular, boundaries are perceptually more important than errors within the body of an image element. The weighting functions generated in the weighting generator 28 are then applied to the output 26 in a visible error calculation unit 29 to produce a "visible error matrix" analogous to that of the audible error matrix described above. The matrix can be defined by a series of coefficients E_{dv1}, E_{dv2}, ..., E_{dvn}. Images are themselves two-dimensional, so for a moving image the visible error matrix will have at least three dimensions.

It should also be noted that the individual coefficients in the audible and visible error matrices may be vector properties.

There are a number of cross-modal effects which can affect the perceived quality of the signal. The effects to be modelled by the cross-modal model 30 may include the quality balance between modalities (vision and audio) and timing effects correlating between the modalities. Such timing effects include sequencing

(event sequences in one modality affecting user sensitivity to events in another) and synchronisation (correlation between events in different modalities).

One key component of the multi-modal model of the present invention is synchronisation. This part of the model is shown in Figure 2. The degree of synchronisation between the inputs is determined in a synchronisation measurement unit 38. This takes inputs from the visual sensory layer (input 38v) and the audible sensory layer (input 38a) relating to the respective delays in the two signals. The synchronisation measurement unit 38 determines the difference in these two delays and generates an output 38v representative of the relative delay between the two signals. This, rather than the absolute delay in either signal, is the perceptually significant. Such lack of synchronisation has been determined in prior art systems but, as will be discussed, the perceptual importance of such synchronisation errors varies according to the nature of the stimulus.

To this end, the cross-modal model 30 also uses information about the audio and video data streams (inputs 35, 36), and optionally the task being undertaken, (input 37) to determine the subjectivity of any synchronisation errors.

In this embodiment the objective parameters describing the audio components of the signals are audio descriptors generated from the input 35 in processor 31. These audio descriptors are RMS energy over a succession of overlapping short intervals of predetermined length, and signal peak and decay parameters. These values give an indication of the general shape and duration of individual audio events.

The parameters describing the video components are video descriptors generated from the input 36 in a processor 32, such as motion vectors, see for example chapter 5 in [*Netravali A N, Haskell B G, "Digital Pictures; representation and compression", Plenum Press, ISBN 0-306-42791-5. June 1991*], and a persistence parameter describing the subjective importance, and the decay of this importance with time.

These parameters are used by a further processor 33 to determine the nature of the content of the stimulus, and generate therefrom a synchronisation error perceptibility value, which is output (39) to the perceptual model 40, along with the actual value of the synchronisation error (output 38s). The perceptual model 40 can then compare the synchronisation error value with the perceptibility

value to generate a perceptual quality value, which contributes to a cross-modal combining function fn_{pm} to be used by the perceptual model 40.

A mathematical structure for the model can be summarised:

$E_{da1}, E_{da2}, \ldots, E_{dan}$ are the audio error descriptors, and

$E_{dv1}, E_{dv2}, \ldots, E_{dvn}$ are the video error descriptors.

Then, for a given task:

fn_{aws} is the weighted function to calculate audio error subjectivity,

fn_{vws} is the weighted function to calculate video error subjectivity, and

fn_{pm} is the cross-modal combining function previously discussed. This function may include other weightings, to account for other cross-modal factors, for example quality mismatches and task-related factors.

The task-specific perceived performance metric, PM, output from the model 40 is then:

$$PM = fn_{pm} [fn_{aws} \{ E_{da1}, E_{da2}, \ldots, E_{dan} \}, fn_{vws} \{ E_{dv1}, E_{dv2}, \ldots, E_{dvn} \}]$$

The perceptual layer model 40 may be configured for a specific task, or may be configurable by additional variable inputs T_{wa}, T_{wv} to the model (inputs 41, 42), indicative of the nature of the task to be carried out, which varies the weightings in the function fn_{pm} according to the task. For example, in a video-conferencing facility, the quality of the audio signal is generally more important than that of the visual signal. However, if the video conference switches from a view of the individuals taking part in the conference to a document to be studied, the visual significance of the image becomes more important, affecting what weighting is appropriate between the visual and auditory elements. These values T_{wa}, T_{wv} may also be fed back to the synchronisation perception measuring function 38, to allow the synchronisation error subjectivity to vary according to the task involved. High level cognitive preconceptions associated with the task, the attention split between modalities, the degree of stress introduced by the task, and the level of experience of the user all have an effect on the subjective perception of quality.

The functions fn_{aws} , fn_{vws} may themselves be made functions of the task weightings, allowing the relative importance of individual coefficients E_{da1}, E_{dv1} etc to be varied according to the task involved giving a prediction of the performance metric, PM' as:

$$PM' = fn'_{pm} [fn'_{aws} \{ E_{da1}, E_{da2}, \ldots, E_{dan}, T_{wa}\}, fn'_{vws} \{ E_{dv1}, E_{dv2}, \ldots, E_{dvn}, T_{wv}\}]$$

A multi-dimensional description of the error subjectivity in the auditory and visual modalities is thereby produced.

In the arrangement of Figure 5 an avatar 29 represented on a screen is generated from an audio input 10, from which the speech content is derived, and an input 20 which provides the basic gesture data for the animation of the avatar in the animation unit 28. The audio input 10 is also supplied to a speaker system 19, and to the animation process. The process requires the selection of the viseme (facial arrangement) appropriate to the sound being uttered (element 27) and this is used to control the animation unit 28. It is desirable to have the visemes synchronised with the sound, but the animation process makes this difficult to achieve. Some visemes are more tolerant of synchronisation errors than others, and so by applying the audio input 10 and the identity of the selected viseme to the synchronisation model 30 (Figure 2) this tolerance can be determined, and used to control the animation process 28, for example by extending or shortening the duration of the more tolerant visemes, to allow better synchronisation of the less-tolerant visemes.

In these embodiments the values derived in the processor 33 depend on the stimulus type. A selection of experimental results showing this inter-relationship is presented below in order to illustrate the influence of stimulus on the perceptual relevance of synchronisation errors.

Figure 3 shows the number of subjects detecting a synchronisation error averaged across three stimulus types. These types are:

(1) an object entering and leaving the field of vision, as an example of a brief visual cue;

(2) an object entering and remaining in the field of vision, as an example of a longer visual cue, and

(3) a speech cue (talking head).

Each visual cue is accompanied by an audible cue generated by the object in the visual cue.

It will be seen from Figure 3 that there is an underlying feature of temporal asymmetry in the perceptibility of synchronisation errors. Synchronisation errors in which the audio signal leads the visual signal are perceptually more important than those in which the visual signal leads the audio signal by the same interval. This is probably because we are used to receiving audio cues later than corresponding visual cues in ordinary experience, since in the natural world the associated physical signals travel at vastly different speeds (340 metres/second for sound and 300 million metres/second for light).

The general form of the results reflects the recommended synchronisation thresholds given in Recommendation J.100 of the ITU, i.e. 20 milliseconds for audio lead and 40 milliseconds for audio lag. This recommendation provides a fixed figure for all content types and is intended to ensure that synchronisation errors remain imperceptible. This approach is suitable for the specification of broadcast systems.

However, it has been found that synchronisation error detection is greater for a long visual cue than for a short visual cue or a visual speech cue. Figure 4 shows the results for these two stimulus types, and for a "talking head", which is a special case because human subjects are highly specialised for speech perception compared with more general content. The two non-speech sound stimuli selected were both relatively abrupt, as these make greater demands on synchronisation than would a continuous noise.

These are shown on a single graph for ease of comparison.

The key features of these results are:

(i) The general trend in error detection asymmetry is apparent for all stimulus types.

(ii) The duration/distinctness of the long ("axe") stimulus, in which the object generating the sound appears, and then remains in view, results in greater probability of error detection than for the shorter ("pen") stimulus, in which the object appears with the sound, but rapidly goes out of view again.

(iii) Error detection for the speech ("Marilyn") stimulus is consistent with the other two stimuli when the audio lags the video, but is greater than for either of the other stimuli when the audio leads the video.

The probability of synchronisation error detection therefore varies with the duration and distinctness of the visual stimulus. Moreover, there is a high sensitivity to synchronisation errors in speech when the audio signal leads the video. This latter result was not expected, since it has been previously argued that during speech perception it is not possible to resolve the timing of "error" events more accurately than the duration of the semantic elements of the speech stream, see for example Chapter 7 in *[Handel, S. "Listening: an introduction to the perception of auditory events", MIT Press, 1989]*. It appears in practice that, perhaps due to the short duration of certain semantic units such as consonant onsets, subjects are very sensitive to audio lead synchronisation errors with talking-head/speech stimuli.

CLAIMS

1. A method of determining the subjective quality of an audio-visual stimulus, comprising the steps of:

measuring the actual synchronisation errors between the audio and visual elements of the stimulus,

identifying characteristics of audio and visual cues in the stimulus,

and generating a measure of subjective quality from said errors and characteristics.

2. A method according to claim 1, wherein the characteristics of the audio and visual cues are used to generate one or more synchronisation error tolerance values.

3. A method as claimed in claim 2, wherein the audio-visual stimulus is monitored for occurrences of synchronisation errors exceeding such tolerance values.

4. A method according to claim 3, wherein the means generating the stimulus is controlled to maintain the synchronisation in a predetermined relationship with the said tolerance values.

5. A method according to claim 4, wherein the resulting measure of subjective quality is used to control the operation of an avatar animation process.

6. Apparatus for determining the subjective quality of an audio-visual stimulus, comprising means for measuring the actual synchronisation errors between the audio and visual elements of the stimulus, means for the identification of characteristics of audio and visual cues in the stimulus, and means for generating a measure of subjective quality from said synchronisation errors and characteristics.

7. Apparatus according to claim 6, wherein the means for identifying cue characteristics generates one or more synchronisation error tolerance values.

8. Apparatus as claimed in claim 7, comprising means for monitoring the audio-visual stimulus for occurrences of synchronisation errors exceeding said tolerance values.

9. Apparatus according to claim 8, comprising means for controlling the means generating the stimulus to maintain the synchronisation in a predetermined relationship with the said tolerance values

10 Apparatus according to claim 9, further comprising animation process means controlled by the subjective quality measurement means to generate an animated image.

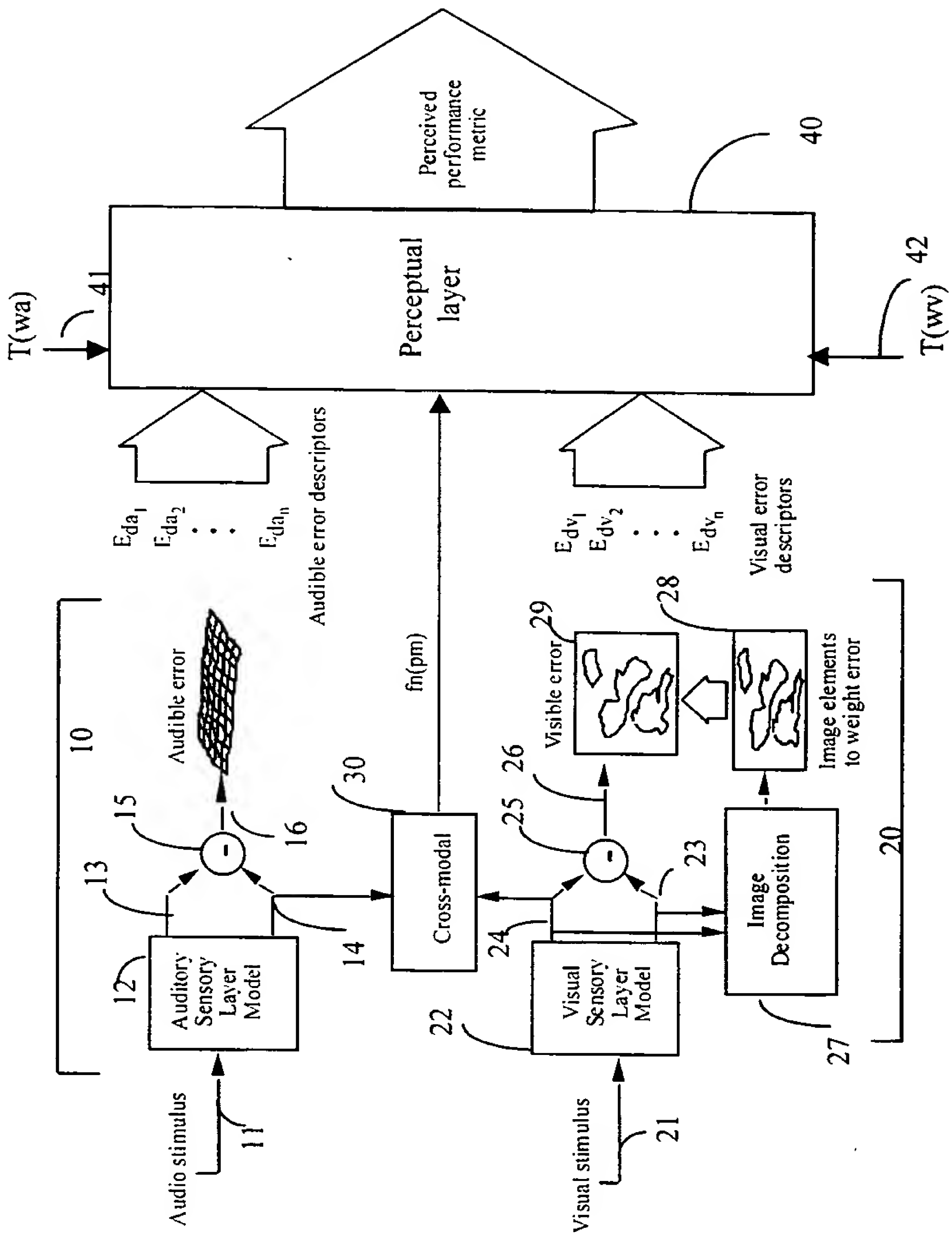

Perceived performance metric
Perceptual layer
40
41
42
T(wa)
T(wv)
Audible error descriptors
Visual error descriptors
fn(pm)
Audible error
Visible error
Image elements to weight error
Auditory Sensory Layer Model
Visual Sensory Layer Model
Cross-modal
Image Decomposition
Audio stimulus
Visual stimulus
10
11
12
13
14
15
16
20
21
22
23
24
25
26
27
28
29
30

Figure 1

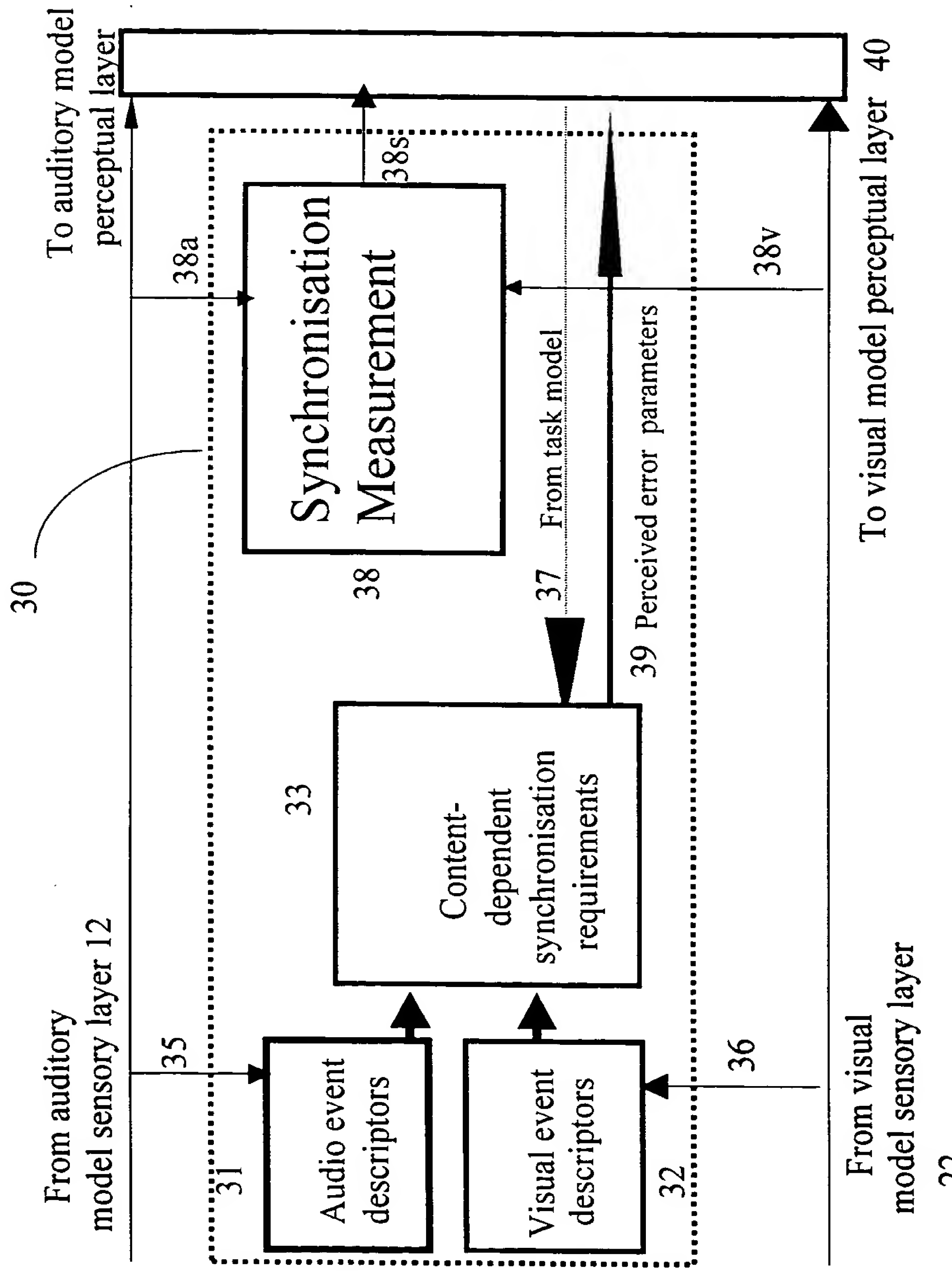

To auditory model perceptual layer
40
To visual model perceptual layer
38s
Synchronisation Measurement
38a
38v
38
From task model
37
39 Perceived error parameters
30
33
Content-dependent synchronisation requirements
From auditory model sensory layer 12
35
Audio event descriptors
31
Visual event descriptors
32
36
From visual model sensory layer 22

Figure 2

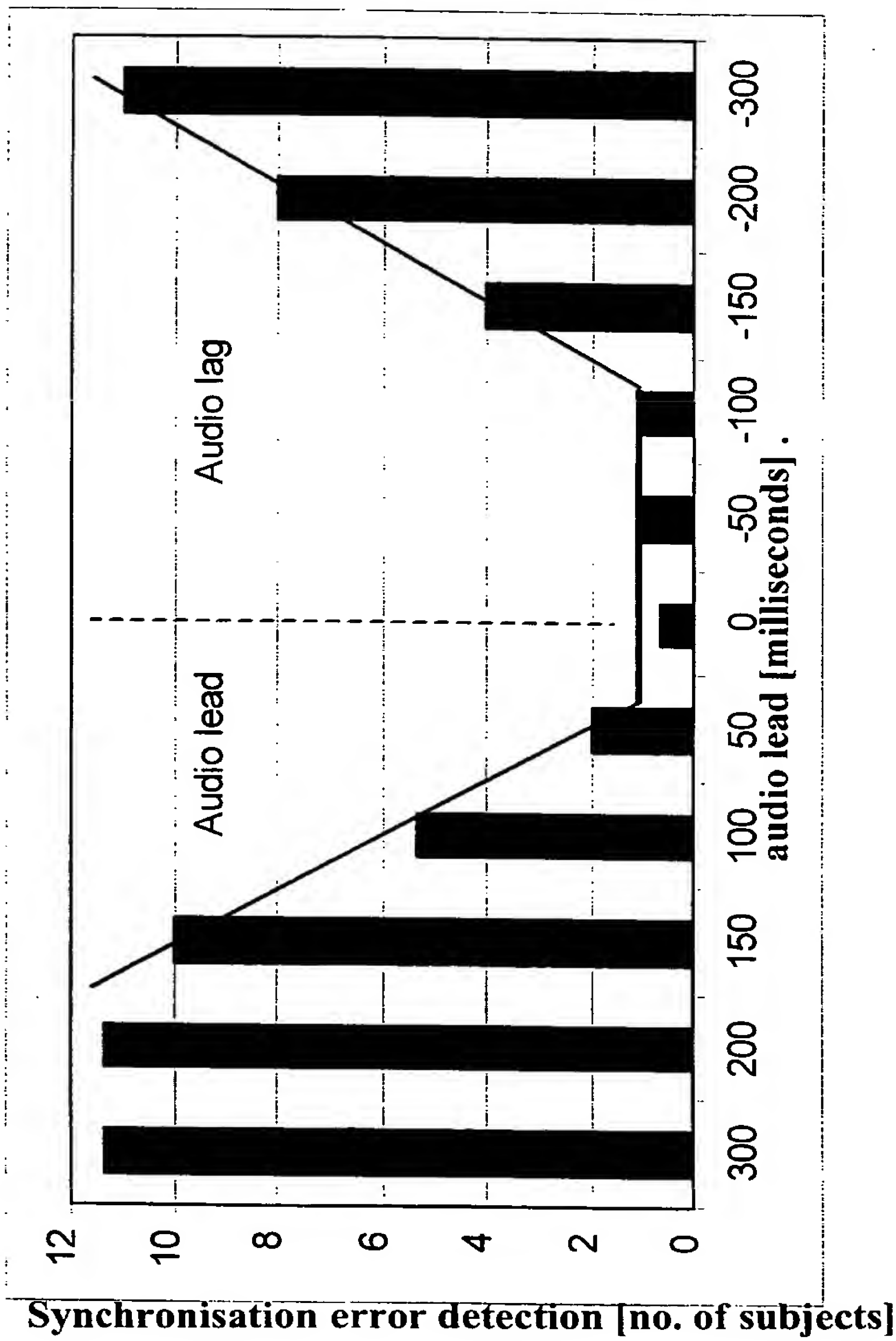

Audio lag
Audio lead
300 200 150 100 50 0 -50 -100 -150 -200 -300
audio lead [milliseconds]
0 2 4 6 8 10 12
Synchronisation error detection [no. of subjects]

Figure 3

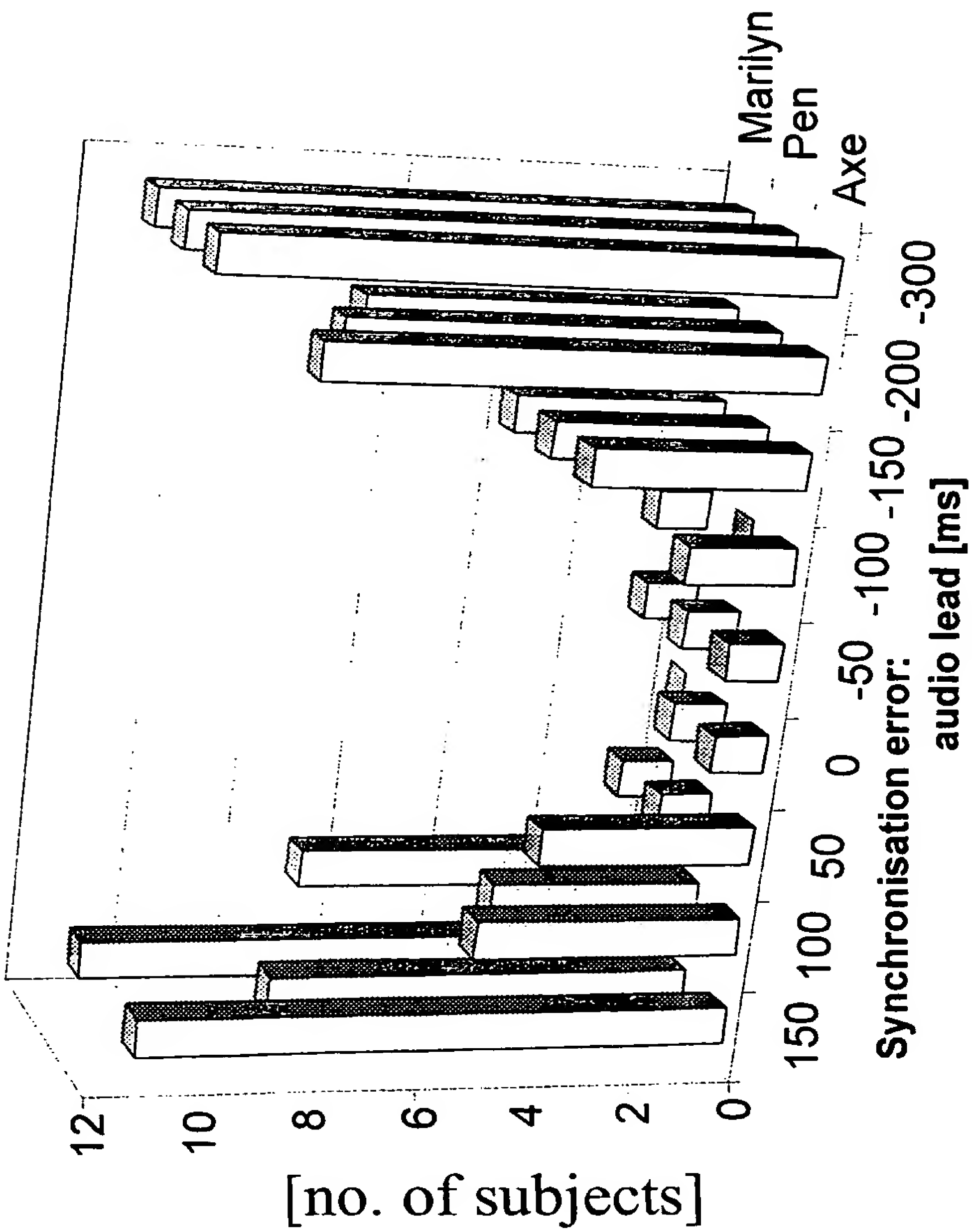

Marilyn
Pen
Axe
-300
-200
-150
-100
-50
0
50
100
150
Synchronisation error:
audio lead [ms]
12
10
8
6
4
2
0
[no. of subjects]

Figure 4

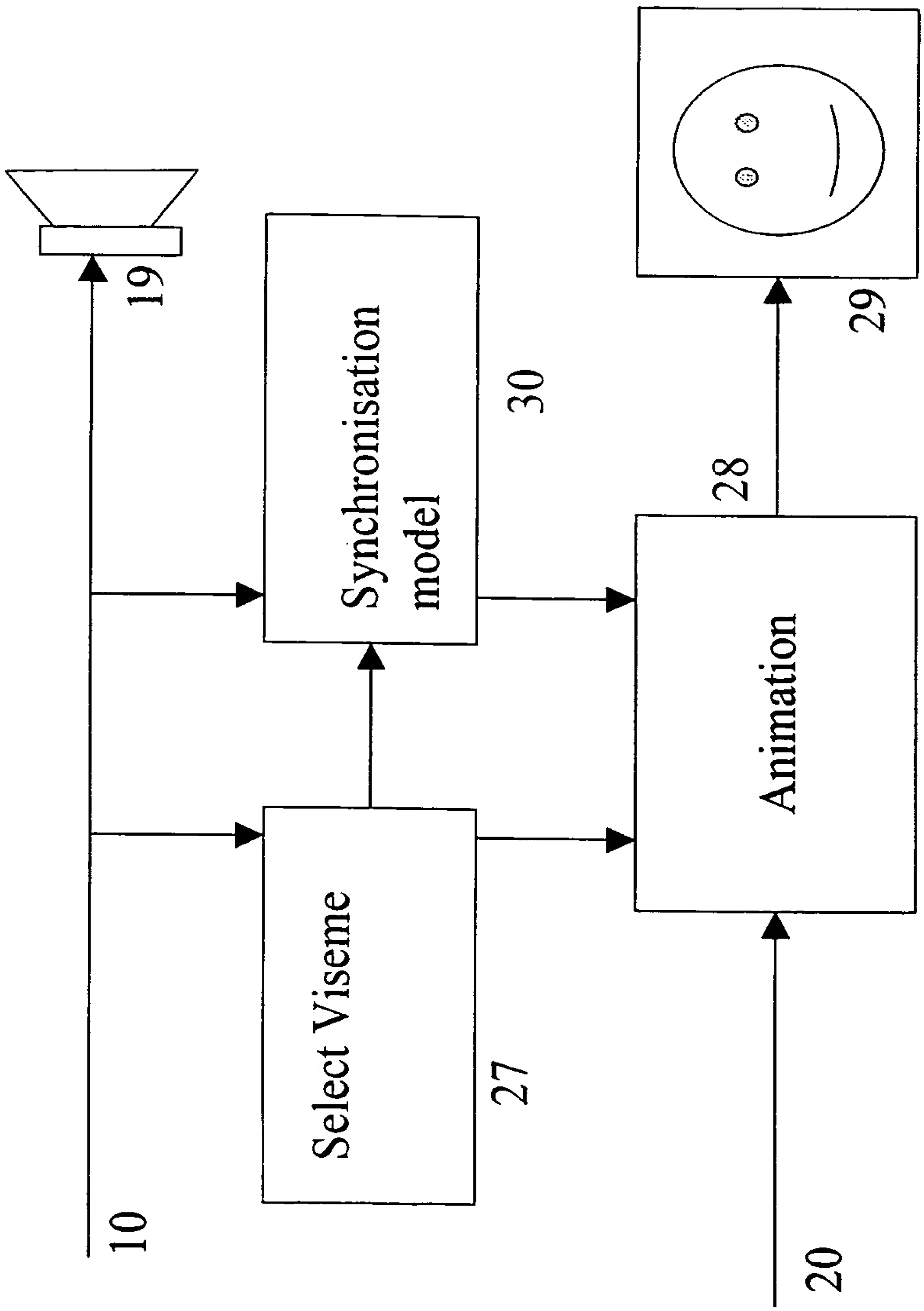

Synchronisation model
Select Viseme
Animation
10
19
20
27
28
29
30

Figure 5

• a scenario-specific perceptual layer 40.

An auditory sensory layer model component 10 comprises an input 11 for the audio stimulus, which is provided to an auditory sensory layer model 12. The auditory model 12 measures the perceptual importance of the various auditory bands and time elements of the stimulus, and generates an output 16 representative of the audible error as a function of auditory band (pitch) and time. This audible error may be derived by comparison of the perceptually modified audio stimulus 13 and a reference signal 14, the difference being determined by a subtraction unit 15 to provide an output 16 in the form of a matrix of subjective error as a function of auditory band and time, defined by a series of coefficients E_{da1}, E_{da2}, ..., E_{dan}. Alternatively the model may produce the output 16 without the use of a reference signal, for example according to the method described in international patent specification number WO96/06496.

A similar process takes place with respect to the visual sensory layer model 20. However, in this context a further step is required. The image generated by the visual sensory layer model 22 is analysed in an image decomposition unit 27 to identify elements in which errors are particularly significant, and weighted accordingly, as described in international patent specification number WO97/32428. This provides a weighting function for those elements of the image which are perceptually the most important. In particular, boundaries are perceptually more important than errors within the body of an image element. The weighting functions generated in the weighting generator 28 are then applied to the output 26 in a visible error calculation unit 29 to produce a "visible error matrix" analogous to that of the audible error matrix described above. The matrix can be defined by a series of coefficients E_{dv1}, E_{dv2}, ..., E_{dvn}. Images are themselves two-dimensional, so for a moving image the visible error matrix will have at least three dimensions.

It should also be noted that the individual coefficients in the audible and visible error matrices may be vector properties.

There are a number of cross-modal effects which can affect the perceived quality of the signal. The effects to be modelled by the cross-modal model 30 may include the quality balance between modalities (vision and audio) and timing effects correlating between the modalities. Such timing effects include sequencing

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference A25650 WO	FOR FURTHER ACTION see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
International application No. PCT/GB 99/ 01894	International filing date *(day/month/year)* 15/06/1999	(Earliest) Priority Date *(day/month/year)* 29/06/1998
Applicant BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 2 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

   [ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

   [ ] contained in the international application in written form.
   [ ] filed together with the international application in computer readable form.
   [ ] furnished subsequently to this Authority in written form.
   [ ] furnished subsequently to this Authority in computer readble form.
   [ ] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
   [ ] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).
3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title**,

   [X] the text is approved as submitted by the applicant.
   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

   [X] the text is approved as submitted by the applicant.
   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 2

   [X] as suggested by the applicant.
   [ ] because the applicant failed to suggest a figure.
   [ ] because this figure better characterizes the invention.

   [ ] None of the figures.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 99/01894

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N17/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 596 364 A (WOLF STEPHEN ET AL) 21 January 1997 (1997-01-21) abstract ---	1,6
A	US 5 572 261 A (COOPER J CARL) 5 November 1996 (1996-11-05) abstract ---	1,6
A	ALPERT T ET AL: "SUBJECTIVE EVALUATION OF MPEG-4 VIDEO CODEC PROPOSALS: METHODOLOGICAL APPROACH AND TEST PROCEDURES" SIGNAL PROCESSING. IMAGE COMMUNICATION, vol. 9, no. 4, May 1997 (1997-05), pages 305-325, XP000700943 paragraph '0002! -----	1,6

[ ] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 September 1999	14/09/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Yvonnet, J

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 99/01894

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5596364 A	21-01-1997	US 5446492 A	29-08-1995
US 5572261 A	05-11-1996	NONE	

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark Office
Box PCT
Washington, D.C.20231
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 02 February 2000 (02.02.00)	
International application No. PCT/GB99/01894	**Applicant's or agent's file reference** A25650 WO
International filing date (day/month/year) 15 June 1999 (15.06.99)	**Priority date** (day/month/year) 29 June 1998 (29.06.98)
Applicant HOLLIER, Michael, Peter	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

20 December 1999 (20.12.99)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was
[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland** Facsimile No.: (41-22) 740.14.35	Authorized officer Olivia RANAIVOJAONA Telephone No.: (41-22) 338.83.38

· 10

5

PATENT COOPERATION TREATY

PCT

REC'D 03 JUL 2000
WIPO PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference A25650 WO	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/GB99/01894	International filing date *(day/month/year)* 15/06/1999	Priority date *(day/month/year)* 29/06/1998
International Patent Classification (IPC) or national classification and IPC H04N17/00		
Applicant BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   ☒ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 1 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☐ Certain documents cited
   - VII ☒ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
20/12/1999	29.06.2000
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer de Dieuleveult, A Telephone No. +49 89 2399 8946

I. Basis of the report

1. This report has been drawn on the basis of (*substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

Description, pages:

1-4,6-10	as originally filed			
5	as received on	14/06/2000	with letter of	14/06/2000

Claims, No.:

1-10	as originally filed

Drawings, sheets:

1/5-5/5	as originally filed

2. The amendments have resulted in the cancellation of:

   - ☐ the description, pages:
   - ☐ the claims, Nos.:
   - ☐ the drawings, sheets:

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

4. Additional observations, if necessary:

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB99/01894

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	2-5,7-10
	No:	Claims	1,6
Inventive step (IS)	Yes:	Claims	2-5,7-10
	No:	Claims	1,6
Industrial applicability (IA)	Yes:	Claims	1-10
	No:	Claims	

2. Citations and explanations

see separate sheet

VII. Certain defects in the international application

The following defects in the form or contents of the international application have been noted:

see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

see separate sheet

V. Reasoned statement

1. Reference is made to the following documents:
   D1: US-A-5 596 364 (WOLF STEPHEN ET AL) 21 January 1997
   D2: US-A-5 572 261 (COOPER J CARL) 5 November 1996

2. Claim 1:
   All the features of this claim can be read onto document D1 (see section VIII below) which discloses (see col. 12, lines 13-64) a method of determining the subjective quality of an audio-visual stimulus, comprising the steps of:
   measuring the actual synchronisation errors between the audio and visual elements of the stimulus ("audio-visual synchronization" 165' on the side of the destination instruments; see also col. 1, lines 27-35 and col. 14, lines 51-56),
   identifying characteristics of audio and visual cues in the stimulus (audio and video features 119 and 9, respectively), and
   generating a measure of subjective quality from said errors and characteristics (see "the quality score 14" throughout the description).

   Besides, document D2 also discloses a method for identifying characteristics (such as "opening" and "closing" of the mouth of a speaker) of audio and visual cues in the stimulus and measuring the actual synchronisation errors between the audio and visual elements of the stimulus (see its abstract).

   Consequently, the claimed subject-matter is not novel with respect to the disclosure of D1, or at least not inventive with regard to the combined disclosure of D1 and D2.

3. Claim 6:
   The same objection likewise applies to this corresponding apparatus claim.

4. Claims 2 and 7:
   There is no suggestion in either prior art reference that a tolerance value for synchronisation error should be calculated.
   These dependent claims therefore appear to meet the requirements of Article 33 PCT.

5. Claims 3-5 and 8-10:
These claims are dependent on claims 2 or 7 and, as such, also satisfy Article 33 PCT.

VII. Certain defects

1. Independent claims 1 and 6 are not in the two-part form in accordance with Rule 6.3(b) PCT, which in the present case would be appropriate, with those features known in combination from the prior art (document D1) being placed in the preamble (Rule 6.3(b)(i) PCT) and with the remaining features being included in the characterising part (Rule 6.3(b)(ii) PCT).

2. The features of the claims are not provided with reference signs placed in parentheses (Rule 6.2(b) PCT).

3. Part of the description seems to be missing on page 6, line 11, after "is the perceptually significant".

4. Furthermore:
- on page 4, line 19: "ore" should read "more";
- on page 6, line 9: "38v" should read "38s".

VIII. Certain observations

The scope of claims 1 and 6 appears to be quite large since there are no precision as to which "characteristics" are to be identified.
Besides, how the "measure of subjective quality" is to be generated from the "synchronisation errors" and said "characteristics" still remains obscure.
These independent claims are thus considered to merely define the result to be achieved and not to include the technical features necessary thereto (Article 6 PCT).